United Nations ECE/CES/2018/14/Add.15

Economic and Social Council

Distr.: General
15 May 2018

English only

Economic Commission for Europe

Conference of European Statisticians
Sixty-sixth plenary session
Geneva, 18-20 June 2018
Item 11 (a) of the provisional agenda
Programme of work of the Statistics subprogramme of the United Nations Economic Commission for Europe
Reports on the work of the Conference of European Statisticians, its Bureau and Teams of Specialists

Implementation of the United Nations Economic Commission for Europe Statistical Programme 2017

Note by the secretariat

Addendum

Report of the joint UNECE/UN-GGIM: Europe Workshop on Integrating Geospatial and Statistical Standards

Summary

The document presents the key outcomes of the joint UNECE/UN-GGIM: Europe Workshop on Integrating Geospatial and Statistical Standards which took place from 6 to 8 November 2017 in Stockholm, Sweden. The meeting was conducted under the work programme of the Conference of European Statisticians.

The report is submitted to the Conference of European Statisticians for information.

GE.18-07811(E)

1807811

Please recycle

I. Attendance

1. Under the work programme of the Conference of European Statisticians, the joint UNECE/UN-GGIM: Europe Workshop on Integrating Geospatial and Statistical Standards was held in Stockholm, Sweden, from 6-8 November 2017. The workshop was aimed at experts on statistical and geospatial metadata or technical standards relevant to the modernisation of statistics, in addition to experts from national mapping, cadastral or environmental agencies.

2. It was attended by participants from: Albania, Australia, Austria, Belgium, Brazil, Canada, Finland, France, Germany, Ireland, Israel, Italy, Japan, Latvia, Mexico, Mongolia, Netherlands, New Zealand, Norway, Philippines, Poland, Portugal, Serbia, Slovenia, Spain, Sweden, Switzerland, the former Yugoslav Republic of Macedonia, Turkey, United Kingdom of Great Britain and Northern Ireland, United States of America, as well as by representatives from the European Commission, EuroGeographics, UN GGIM, UNECA, UNECE, UNMIK, and OECD. Participants from research institutes, standards organisations, universities and private sector also attended the workshop.

II. Organization of the meeting

3. The workshop was conducted through a combination of presentations by experts from the national/regional statistical institutes and geospatial information agencies, panel and group discussion sessions. The programme of the workshop consisted of the following substantive topics.

- Session 1: Introduction to statistical and geospatial standards and models;
- Session 2: Examples of integrating statistical and geospatial standards and models;
- Session 3: Challenges and solutions for creating geospatial statistical outputs;
- Session 4 (panel): Future work relevant to statistical and geospatial standards.

4. Ms. Therese Lalor (UNECE) opened the workshop and Mr. Janusz Dygaszewicz, chair of the workshop from the Central Statistical Office of Poland, welcomed participants.

5. Ms. Helen Stoye, Acting Director General of Statistics Sweden, welcomed the participants on behalf of Statistics Sweden. She noted that users increasingly need digitized services combining statistical and geospatial information. The workshop provides a unique opportunity to increase the collaboration between the statistical and geospatial communities.

6. The following persons acted as Session Organizers:

- Session 1: Klas Blomqvist (Swedish Research Council) and Luis Gerardo Esparza (National Institute of Statistics and Geography, Mexico);
- Session 2: Marie Haldorson (Statistics Sweden) and Janusz Dygaszewicz (Central Statistical Office, Poland);
- Session 3: Timothy Trainor (U.S. Census Bureau) and Ekkehard Petri (Eurostat);
- Session 4: Martin Brady (Australian Bureau of Statistics) and Carol Agius (EuroGeographics).

7. All background documents, presentations and the final report for the meeting are available on the website of the UNECE Statistical Division: https://www.unece.org/index.php?id=45404

III. Summary of the discussion and main conclusions reached at the meeting

A. Introduction to Statistical and Geospatial Standards and Models

8. This session was chaired by Klas Blomqvist (Swedish Research Council) and Luis Gerardo Esparza (National Institute of Statistics and Geography, Mexico).

9. The session opened with a panel discussing the statistical standards (Generic Activity Model for Statistical Organizations, Generic Statistical Business Process Model, Generic Statistical Information Model and Common Statistical Production Architecture). The discussion raised the following points:

- The models are generic and are not prescriptive. There is flexibility to allow for national implementations. For example, Brazil has developed a geospatial process model based on Generic Statistical Business Process Model (GSBPM);
- The Generic Statistical Information Model (GSIM) does not include unstructured data, but does include referential metadata such as quality reports;
- The strength of the presented models is that they facilitate the statistical community to develop rigorous systems. These types of models are not necessarily a feature of the geospatial community;
- Continuous improvement can occur through the Generic Activity Model for Statistical Organisations (GAMSO) in the 'Capability Improvement' activity. Evaluation activities are present is many places in the model;
- The models have been used in some countries to discuss the business of official statistics with other government departments.

10. The second panel in the session introduced the geospatial frameworks and standards (Global Statistical Geospatial Framework, Open Geospatial Consortium and International Organization for Standardization – Technical Committee 211 Geograpahic Information/Geomatics). The discussion raised the following points:

- The GSGF is a framework, and not a standard. It was developed in partnership between the statistical and geospatial community, and is more directly applicable to the statistical community;
- There are more than 75 standards in the geospatial community. They are all currently used;
- Data exchange is one of the biggest challenges. Interactions with information are changing as the technological environment changes. It is an appropriate time to discuss this issue as changes are needed;
- Statisticians want a comprehensive view of available high quality geospatial data. There are use cases (for example sample design) where geospatial data could be used, but it is not possible to access the data. This will be important for the 2020 census round and beyond;
- New developments such as evolution of satellite data highlights the importance of having interoperable standards to ensure that two communities can address the new user needs more efficiently and seamlessly.

11. Following the panels, three presentations were given:

- United Kingdom - Removing spatial silos: Delivering spatial statistics in the UK through standards;
- Australia - Exploring the role and application of Discrete Global Grid Systems to integrate statistical and geospatial information;
- European Commission - StatDCAT and GeoDCAT application profile.

12. The following points were raised in the discussion:

- Many countries are experiencing similar problems in terms of raising awareness and changing culture. Opportunities to train and increase exposure between the communities will help them to better understand each other;
- The Discrete Global Grid Systems will be developed in an adaptive way. It will be possible to map to whatever level that is required. There is no preference for a particular grid yet but several implementations are looking at consistent grid specifications and implementations. It is important that the different methods interface with each other;
- If all locations have a unique identifier, this can create additional burden when producing linked data. Despite this burden there is value in creating this linked data. Organizations will determine the level of detail that makes sense to them;
- For some grid cells/pixels, the data included may be sensitive (due to the number of statistical units or the underlying unit record data included) and methods to protect these data will be required.

13. The chairs of the session made the following concluding remarks:

- The statistical and geospatial 'standards' are operating on different levels. The models and frameworks of the statistical community are at a conceptual level. The geospatial standards are at a technical level;
- It is important to build understanding between the communities and to identify where the touchpoints are to help establish common projects;
- The demand for information is going to be bigger than any one organisation. The two communities can meet and learn from each other.

B. Ensuring collaboration across communities

14. This session was organized by Carol Agius (EuroGeographics) and Martin Brady (Australian Bureau of Statistics).

15. The first panel session included Chris Body (ISO/TC211), Klas Blomqvist (Swedish Research Council), Pier-Giorgio Zaccheddu (Fed. Agency for Cartography and Geodesy, Germany), Janusz Dygaszewicz (Central Statistical Office, Poland) and Greg Scott (UN-GGIM). The following points were raised in the discussion:

- The geospatial and statistical communities have cooperated for many years. It is important to recognise that in the context of recent growth in understanding, coordination and coherence;
- Both statistical and geospatial organisations provide information for better decision making. There are shared objectives and mandates;
- It would be useful to understand the interdependencies between the communities – what are the gaps and overlaps. Both current and future activities should be considered when discussing this;

- International collaboration has real benefits. It should be supported through national institutional arrangements to ensure that the cooperation can continue to evolve. To collaborate effectively, there needs to be support from senior management and followed up with a long term, sustainable commitment;
- It is important to be clear of the business requirements for collaboration. Common projects and use cases should be identified. The test beds in OGC can be used to test the use cases;
- Sustainable Development Goals are a key driver and opportunity for the communities to collaborate;
- We should not be afraid to broaden the communities. There is always work to do – we need as many people as possible! Possible collaborators at the expert level include industry, statistical organisations, government ministries, mapping agencies, INSPIRE, Earth Observation Community, ISO liaisons, and UN Expert Group on Classifications;
- Two communities are facing information demand that cannot be addressed by one community alone with increasing challenge from outside competitors;
- Data should be fit for many purposes. Data (either geospatial or statistical) are now consumed by many different communities with different requirements;
- Areas for collaboration should be drawn from user needs. How can the communities work together to provide answers to user questions?
- The word standard needs to be clarified. Many of the statistical 'standards' are frameworks and models, and some of the technical geographic standards could also be called classifications.

C. Examples of Integrating Statistical and Geospatial Standards and Models

16. This session was organized by Marie Haldorson (Statistics Sweden) and Janusz Dygaszewicz (Central Statistical Office, Poland). The session was opened by Marie Haldorson who reflected that we need to be clear what is meant by models, frameworks and standards. Models and frameworks are generic and are open to many different contexts and provide countries with something to benchmark against. On the other hand, standards are more precise, defined and cannot be too generic so that they may be interpreted in a specific way without risk of misunderstanding.

17. The session included the following presentations:

- UNECA - Strategic framework for the integration of statistical and geospatial information in Africa;
- European Commission - Integration of INSPIRE & SDMX data infrastructures;
- Mexico - Mexico's experience in integrating geographic and statistical information;
- Germany - Integrating GSGF and GSBPM – Experiences from Official Statistics Germany;
- Australia - Location information in statistical modernisation transformation;
- Finland - Towards connecting geospatial information and statistical standards in statistical production: Two cases from Statistics Finland;

- Italy - The Experience of reusing WebGIS application templates applied to the integration of statistical and geospatial information;
- Poland - Harmonization of statistical and geodetic divisions in the context of 10 level model, to develop a common framework as a standard of geospatial data production;
- Mongolia - Integrating satellite imagery and geo-spatial data with administrative registration to produce official statistics;
- Switzerland - The collaborative approach between the Federal Statistical Office and the Federal Office of Topography for maintaining a geocoded building and dwelling register used as a base for the production of geostatistical data;

18. The following points were raised in the discussion:

- It is difficult to quantify the benefits of INSPIRE. There are many benefits reported relating to the opening up of public data;
- The Statistical Data and Metadata eXchange (SDMX) Data Structure Definitions (DSDs) for Census are INSPIRE compliant. Also, SDMX and GSIM have been mapped, so it should be possible to map geospatial data to GSIM. This should be examined in the GSIM review;
- The GSIM mapping to SDMX should be updated following the revision;
- Using different geometric layers can open new possibilities for reporting data. For example, a line can be used to represent statistical indicators on roads and transport;
- Both statistical and geographic agencies store unit level data and only disseminate data at aggregate levels. It is particularly important to store georeferenced statistical and administrative data at as fine a level as possible, ideally at the point level, as geographic areas that data is aggregated to can change over time;
- Most countries have a unique identifier for an address. Other countries have a unique identifier for a household. It is possible to have both.

19. Following the presentations, the participants formed small groups to discuss future collaboration opportunities. The groups proposed several opportunities:

- Provide input into the GSBPM and GSIM reviews on how geospatial processes and information can be represented;
- Encourage statistical offices to consider geospatial data as part of their core data;
- Include statistical data in the UNGGIM fundamental data;
- Identify common projects that are centered on the needs of users and clients;
- Collaborate on address and building registers, which are a key need for both communities. One suggestion is to work on an ontology for addresses and buildings that could be used as a reference;
- Initiate a project on data exchange – how to connect the pipes between statistical and geospatial offices;
- Keep communicating and clarifying the language between the communities. It is important to raise awareness and change perspectives;
- Develop communication materials explaining the different frameworks and how they work together;
- Involve vendors who can build tools to support standards;

- Document the common core metadata for geospatial data;
- Identify the key people in both communities who can make decisions and lead the way;
- Consider opportunities in relation to SDGs and Census 2020.

D. Challenges and Solutions for Creating Geospatial Statistical Outputs

20. This session was organized by Timothy Trainor (U.S. Census Bureau) and Ekkehard Petri (Eurostat). It included the following presentations:

- Canada – Canada's approach to integrating socio-economic and environmental statistics with geo-spatial information;
- Poland – Investigation of linked open data technologies for purposes of publishing georeferenced statistical data;
- Japan – World Grid Square Statistics and their application to data analytics;
- Portugal – Location analytics in administrative data to produce House Price Statistics in Portugal;
- United States – Innovative effort to transform and expand dissemination of Census Bureau content to improve service to internal and external customers;
- Finland – OCG Table Joining Service standard revision;
- Institute for Employment Research - Enabling spatial research using German administrative data – A grid-cell approach.

21. The following points were raised in the discussion:

- Implementation experience has shown that statistical data needed to be adjusted to meet the geography standards on open platforms. It is not trivial work, but it can be done.
- Governments value data today in ways they have not in the past. This environment can change the risk appetite of organisations who hold data. There is a move in some countries to be less risk adverse regarding access to data;
- Data from statistical offices that are put on open platforms are subject to the usual confidentiality processes used for any output;
- In the implementations discussed, often the statistical data is stored separately from the geospatial data. This is often due to organisational issues. While there are some distinct aspects to geospatial information these should not be considered special (unusual) and therefore the reason for separate storage should be questioned;
- One lesson learned from implementation experience is to not overcomplicate it before you start;
- The use of standards in implementations can be successful, but the standardisation of the human factor is often difficult.

E. Future Work relevant to Statistical and Geospatial Standards

22. This session was organized by Carol Agius (EuroGeographics) and Martin Brady (Australian Bureau of Statistics).

23. The second panel session included Ingo Simonis (OGC), Márta Nagy-Rothengass (Eurostat), Timo Aarnio (National Land Survey of Finland), Ian Coady (Office for National Statistics, UK), Tim Trainor (U.S. Census Bureau) and David Barraclough (OECD).

24. The following challenges and opportunities were raised in the discussion:

- Concepts - We need agreement on conceptual issues, so they can be reflected in the standards. Technology can bring conceptual models together;
- Standards – When exchanging data in a machine to machine manner, standards are important. The geospatial dimension is missing in SDMX, this should be discussed;
- Data - Integration is not rocket science. It is important to ensure that you are combining the correct values. Having persistent identifiers would help with this;
- Culture - There is a need to build understanding – not just between statistical and geospatial communities, but also between people who work in standards governance and technical people;
- Communication - It is not always clear how to find information on the standards, how they relate to each other and how to begin using them;
- Management – It is important to raise the importance of this work with the senior management. It is important to build high level support. Census 2020 and the SDGs are a key driver, so benefits can be linked to them;
- Timeliness – The world is changing, and we may need new standards. The standards development process is not necessarily a fast process;
- Working method – the communities should identify concrete projects and use cases. These should be worked in an iterative or agile way. Use small pilot projects to show the value of the work.

25. The following proposals were made for future work:

- A pitch statement to present to senior managers and leaders to gain buy-in and funding;
- A beginner's guide to using standards from both communities;
- Both communities to develop communication materials that simply describe the interrelationships between their frameworks, models and standards;
- Pilot to determine options for persistent ids to link aggregate statistical outputs to standard geographies;
- Look for opportunities to work on semantic interoperability issues (for example, ontology for addresses and buildings);
- Improve the discoverability of geospatial tools that are based on standards;
- Guidance on how to store geospatial objects references/links in existing statistical databases;
- Map the data exchange process between statistical and geospatial organisations
- Examine comparative use cases for application of traditional geography and emerging grid technologies, particularly for dissemination of statistics;
- Examine pathways and interest within Statistical Community to move to formal ISO Standards for models and frameworks in addition to existing ones (e.g. ISO/TC 154);